Supplementary Information:

Neural correlates of perisaccadic visual mislocalization in extrastriate cortex

Geyu Weng[1,2], Amir Akbarian[2], Kelsey Clark[2], Behrad Noudoost[2,*], Neda Nategh[2,3,*]

[1]Department of Biomedical Engineering, University of Utah, Salt Lake City, UT, USA. [2]Department of Ophthalmology and Visual Sciences, University of Utah, Salt Lake City, UT, USA. [3]Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT, USA

*** Correspondence**

Behrad Noudoost

behrad.noudoost@utah.edu

Neda Nategh

neda.nategh@utah.edu

Word counts

1740

Number of figures

6 + 1 table

Supplementary Methods

Empirical RF estimation

For each probe location, the probe-aligned responses were calculated by averaging the spike trains over repetitions of the probe before or after the saccade (greater than 100 ms before or after the saccade onset), from 0:200 ms following probe presentation, across all trials, which provided an estimate of the temporal RFs. The response was then smoothed using a Gaussian window of 5-ms full width at half maximum. The average of these stimulus-aligned responses around the neuron's response latency (ranges from ~50 to ~120 ms), measured for all probes, was used to estimate the spatial RF for each neuron. The empirical spatial and temporal RFs can be obtained for large enough time windows over times to saccade onset.

Dimensionality reduction

The fast dynamics of the neurons' spatiotemporal sensitivity around saccades requires a high-dimensional representation of STUs. We designed the behavioral paradigm with probe stimuli appearing every 7 ms at random locations over a 9×9 grid, which enabled us to create a discretized representation of the neuron's spatiotemporal sensitivity. First, by binning the time and delay dimensions (7 ms bins) we could reduce the dimensionality by about two orders of magnitude, which was still not feasible for a computationally robust estimation. Next, we designed a set of temporal basis functions $\mathcal{B}_{i,j}(t,\tau)$ using second order B-spline functions $\mathcal{U}_i(\tau)$ and $\mathcal{V}_j(t)$ to parameterize the kernels corresponding to each location across the time t and delay τ dimensions as follows:

$$\mathcal{B}_{i,j}(t,\tau) = \mathcal{U}_i(\tau)\mathcal{V}_j(t) \quad (1)$$

where $\{\mathcal{U}_i(\tau)\}$ spans across delay τ and represents a kernel that lasts for 200 ms, constructed by a set of 33 evenly spaced knots at {-13, 6, …, 204, 211} ms, which correspond to a total of 30

basis functions. Similarly, $\{\mathcal{V}_j(t)\}$ spans across time t and represents a saccade-aligned kernel that lasts for 1081 ms, constructed by a set of 159 evenly spaced knots at {-554, -547, …, 545, 552} milliseconds, which correspond to a total of 156 basis functions. The arrangement of the knots follows the 7-ms binning originally applied to the time and delay dimensions. Using these basis functions, the stimulus kernels in Eq. (1) in Methods section were described as the weighted combination of STUs defined over time, delay, and location, as follows:

$$k_{x,y}(t,\tau) = \sum_{i,j} \kappa_{x,y,i,j} \cdot \mathcal{B}_{i,j}(t,\tau) \quad (2)$$

where $\{\kappa_{x,y,i,j}\}$ represent the weights of the STUs associated with location (x, y), and time and delay bin indices $\{i, j\}$.

The STU weights are estimated by fitting the encoding model (defined in Eq. (1) in Methods), to the spiking data collected from each neuron. However, to further limit the amount of STUs required for the model fitting, we used a statistical significance measure to identify those STUs that had a significant impact on a neuron's response at a specific time[1]. We obtained the weight distribution of each STU, by fitting a simpler generalized linear model (GLM) composed of individual STUs on 100 subsets of randomly selected spike trains (35% of all trials), and compared it to a control distribution obtained from 100 subsets of shuffled trials where the relationship between stimulus and response was altered. The conditional intensity function (CIF)[2,3] of this GLM was defined as:

$$\lambda_t = f(\sum_{\tau=1}^{T} s_{x,y}(t-\tau) \cdot \kappa_{x,y,i,j} \cdot \mathcal{B}_{i,j}(t,\tau)) \quad (3)$$

where λ_t is the instantaneous firing rate of the neuron at time t, $s_{x,y}(\tau)$ is the stimulus history of length T at location (x, y), $\{\kappa_{x,y,i,j}\}$ is the weight of the STU represented by the basis function $\mathcal{B}_{i,j}(t,\tau)$ defined in Eq. (1), and f is a static nonlinearity.

The significance of the contribution of a STU represented by its weight parameter was evaluated by satisfying the following condition:

$$|\mu - \tilde{\mu}| \geq 1.5\tilde{\sigma} \quad (4)$$

which denotes that the absolute difference between the mean of the weight distribution μ—obtained using fitting the above GLM model to original spiking responses, and the mean of the control weight distribution $\tilde{\mu}$—obtained from the shuffled spike trains, should be greater or equal to 1.5 times the standard deviation of the control weights distribution ($\tilde{\sigma}$). The threshold of 1.5 was chosen heuristically to reduce the dimensionality of STU space to ~10^4, making the model fitting process practical without overfitting. Next, we used this subset of STUs to parameterize the linear filtering stage of the SVGLM, characterized by the kernels in Eq. (2) above, and according to the CIF defined in Eq. (1) in Methods section, in a less complex space, with the aim of determining how these STUs build up the neuron's spatiotemporal sensitivity. Note that the sum in Eq. (2) is limited to the subset of $\{x, y, i, j\}$ whose corresponding STU was considered significant based on Eq. (4), and the weights for the remaining STUs were assigned a value of zero.

Model performance

The data were randomly split into a training set (35%), a validation set (30%), and a testing set (35%), and we used the testing set to evaluate the model performance. Supplementary figure 2b shows that the model predicts the neural response at the single trial level. The “good” trial was defined as the trial with the largest normalized log-likelihood difference (ΔLL/spike), and the “average” trial has the median normalized ΔLL/spike. The normalized ΔLL/spike was calculated as the log-likelihood (LL) of the spike trains using the model-predicted firing rate minus that under a null model and normalized by spike counts. The null model is a model where the instantaneous firing rate of the neuron is set to its average firing rate. Supplementary figure 2c compares the normalized ΔLL/spike for fixation (-300:-150 ms) vs. perisaccadic (0:150 ms) time windows and shows that the performance of model-predictions in the perisaccadic period is slightly better than in fixation period (fixation = 0.15±0.00 bits/spike, perisaccadic = 0.16±0.00 bits/spike, p = 0.00), indicating that the model is successfully capturing changes in neural sensitivity around the time

of saccades. Supplementary figure 2d shows the correlation coefficient (CC) between the model-predicted firing rate and the empirical firing rate in response to the repeated presentation of a sequence of probe stimuli falls within the level of the inherent trial-by-trial variability. The data-data CC was measured between binned firing rates in response to the same 300 ms stimulus sequence; data were randomly split (60%-40%) 15 times and the mean is reported. The average firing rate was computed by binning the probe-aligned spikes using non-overlapping windows of 30 ms and smoothing the binned response with a Gaussian window of 5 ms (full width half max) and normalizing to have a mean of zero and unit standard deviation. The data-data CC is significantly higher than the model-data CC (data-data = 0.44±0.01, model-data = 0.30±0.01, p = 5.96e-75).

Analyzing perisaccadic localization by decoding neural responses

To directly evaluate how the neural response can generate a readout of location information, we established a maximum a posteriori decoder based on every neuron's spiking activity. For each neuron, the trials were randomly split into a training set (80%) and a testing set (20%), and we used the testing set to evaluate the decoding performance. The decoded stimulus corresponds to the maximum posterior probability $P(x|r)$, which determines the probe stimulus x that is most likely to have caused the observed firing rate r. The posterior probability is calculated according to Bayes' rule,

$$P(x|r) = \frac{P(r|x)P(x)}{P(r)} \quad (5)$$

where $P(r|x)$ denotes the likelihood of observing the mean firing rate r in response to a stimulus appearing at certain probe location, x, during 50:100 ms after stimulus onset, and $P(x)$ and $P(r)$ denote respectively the prior probability of the stimulus and response, calculated as the percentage of occurrence of probe x and mean firing rate r. The decoder was given firing rates corresponding to a center probe and its 8 neighboring probes, and the decoding was repeated for

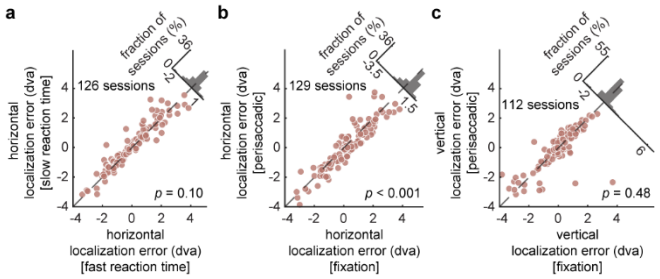
the 6 probes around the ST. For easier visualization, we divided the 9 probes into 3 columns, near (closest to FP), middle, and far (farthest from the FP). The percentage of times the decoded location was in the direction of the saccade or the direction opposite the saccade was compared during presaccadic (-100:0 ms) and perisaccadic (50:150 ms) time windows (Supplementary Fig. 4). 'Incorrectly decoded in the direction of the saccade' means the average percentage for the following three conditions: when the middle probe was presented and the far probe was decoded, when the near probe was presented and the middle probe was decoded, or when the near probe was presented and the far probe was decoded. Similarly, the three conditions for 'incorrectly decoded in the direction opposite the saccade' are: when the middle probe was presented and the near probe was decoded, when the far probe was presented and the middle probe was decoded, or when the far probe was presented and the near probe was decoded. The decoder was more likely to be confused in the saccade direction presaccadically compared to perisaccadically (p = 1.21e-12), but more likely to be confused in the direction opposite the saccade perisaccadically compared to presaccadically (p = 9.12e-06). In other words, neural responses to each probe became more similar to those located closer to the FP perisaccadically, and thus more difficult for the decoder to differentiate, leading to a confusion in the readout location opposite to the saccade direction, similar to what we observed using the spatial bias measure (Fig. 3d).

Comparing responses and bias-relevant STU maps between MT vs. V4 neurons

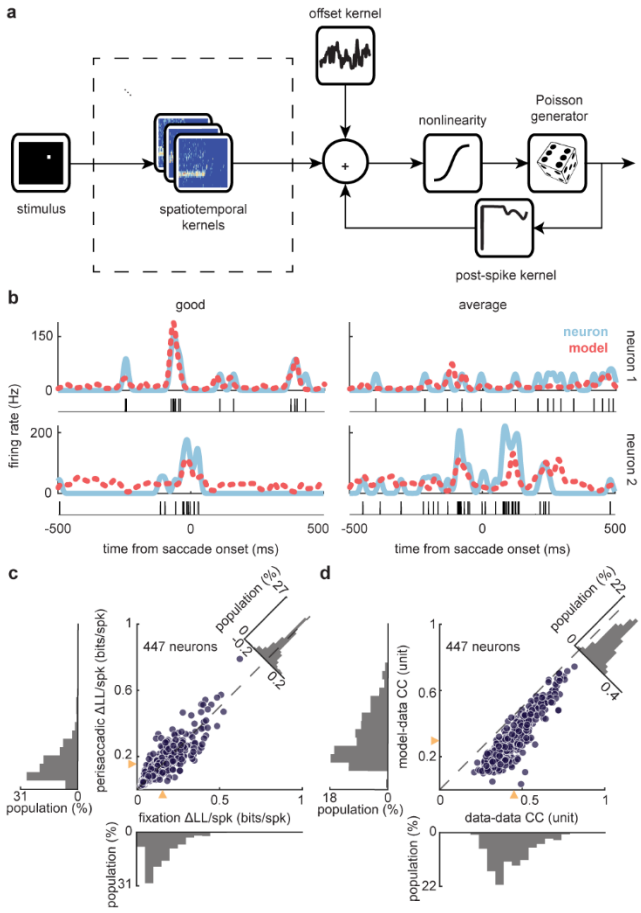
Supplementary figure 5 shows the model response for MT vs. V4 neurons for neurons with d < 11 and $d \geq 11$. For MT neurons with $d < 11$, we outlined the regions corresponding to bias-relevant STUs with 60% contour, which is around time from stimulus onset 70:110 ms and time of stimulus from saccade onset -20:10 ms (top left). V4 neurons with $d < 11$ was outlined with 50% contour and show bias-relevant response at a slightly earlier time from stimulus onset 60:100 ms and time of stimulus from saccade onset -20:10 ms (top right). MT neurons with $d \geq 11$ have two regions

of bias-relevant response at 52% contour (bottom left). The first region is around time from stimulus onset 60:100 ms and time of stimulus from saccade onset 0:20 ms, and the second region is around time from stimulus onset 40:100 ms and time of stimulus from saccade onset 40:230 ms. Similarly, V4 neurons with $d \geq 11$ was outlined with 56% contour, and the first region of bias-relevant response is around time from stimulus onset 50:100 ms and time of stimulus from saccade onset 0:20 ms, and the second region is around time from stimulus onset 40:100 ms and time of stimulus from saccade onset 70:240 ms (bottom right).

Supplementary Figures

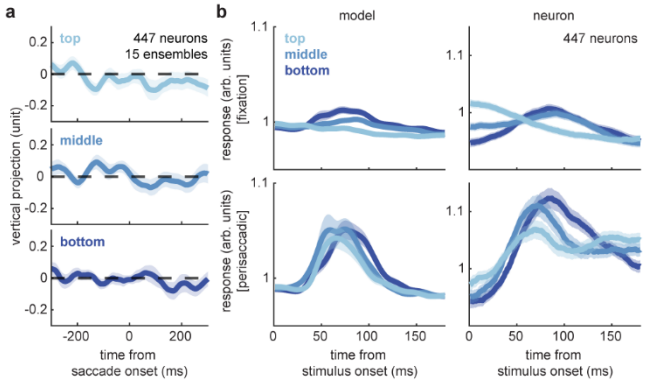


Supplementary Fig. 1. Comparing localization errors based on reaction time, and horizontal vs. vertical directions. a. Horizontal localization error for each session, with trials divided by median reaction time (x-axis, fast vs. y-axis, slow). Plot shows horizontal localization error for probes appearing perisaccadically (-100:0 ms), combining sessions for both Monkey 1 and Monkey 2 (p = 0.10). **b**. Horizontal localization error for each of 110 sessions, averaged across probes near the ST, presented in the fixation (x-axis; -250:-150 ms) vs. perisaccadic (y-axis; -100:0 ms) windows (p = 3.68e-05). **c.** Vertical localization error of each of 110 sessions for probes presented in the fixation (x-axis; -150:-100 ms) vs. perisaccadic (y-axis; -100:0 ms) windows (p = 0.48). We only measured the vertical localization errors for probes below the ST to avoid averaging out potential compression effects. Histograms in the upper right of all panels show the distribution of differences.

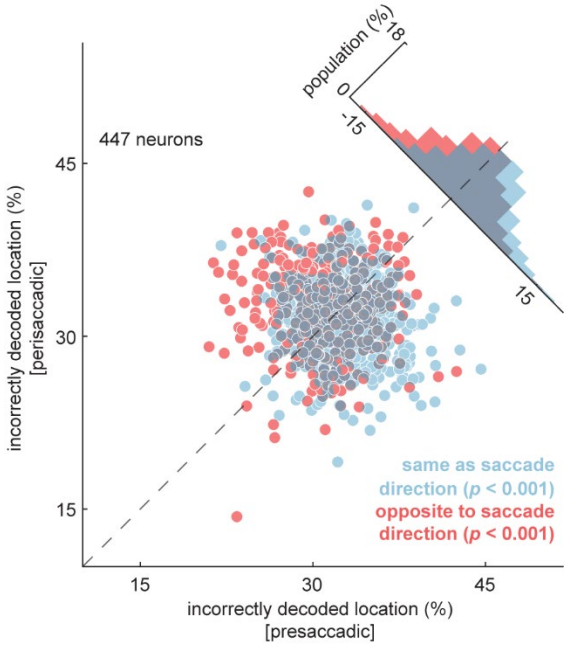


Supplementary Fig. 2. Structure and performance of the SVGLM. a. Schematic of the SVGLM. The stimulus is convolved with 4-dimensional kernels (consisting of STUs) representing the time-varying spatiotemporal sensitivity of individual neurons. The filter stimulus is then added to the output of an offset kernel and the signal generated by a post-spike kernel. The sum passes through a nonlinearity to estimate the neuron's spike rate which is used as the underlying rate by a Poisson spike generator to predict the neuron's spiking activity. **b.** The recorded neural response vs. model-predicted response for a "good" trial (best ΔLL/spike) and an "average" trial (median ΔLL/spike) of two example neurons. Spikes in each trial are shown below the smoothed traces. **c.** Comparison of the normalized ΔLL/spike of the recorded spikes under the model-predicted response in fixation (-300:-150 ms) vs. perisaccadic (0:150 ms) windows. Yellow

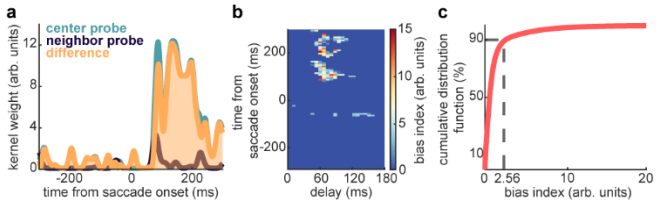
triangles illustrate the medians (fixation = 0.15, perisaccadic = 0.16); histograms show the marginal distributions (left, bottom) and the difference distribution (upper right). **d.** Comparison of the normalized CC between the data-data correlation vs. model-data correlation, evaluating variability in responses to the same stimulus. Yellow triangles illustrate the medians (data-data = 0.45, model-data = 0.30); histograms show the marginal distributions (left, bottom) and the difference distribution (upper right).



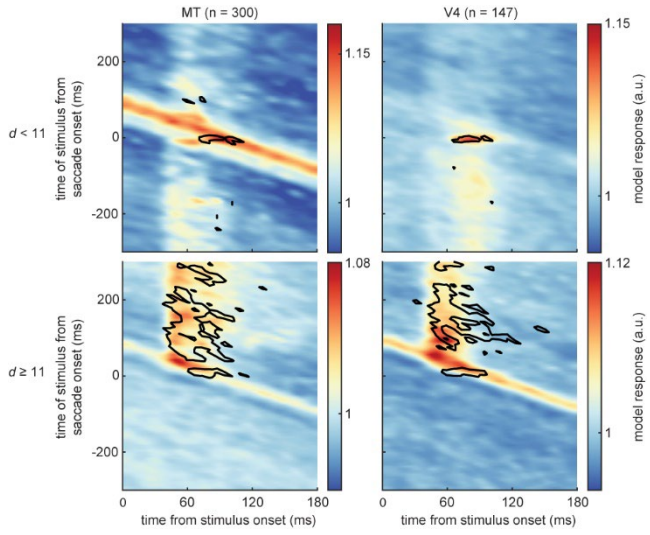
Supplementary Fig. 3. Vertical spatial bias computed using the neurons' kernels. a. Mean vertical spatial bias over time from saccade onset of 15 ensembles of neurons, over delay 50:100 ms relative to each timepoint (x-axis). The vertical spatial bias is the same during fixation (-500:-300 ms) and perisaccadically (50:150 ms) for top, middle, and bottom probes with respect to the ST location (top: fixation = -0.01±0.03, perisaccadic = -0.08±0.06; p = 0.11; middle: fixation = -0.00±0.04, perisaccadic = -0.04±0.05; p = 0.23; bottom: fixation = -0.01±0.03, perisaccadic = -0.00±0.05; p = 0.72). Shaded area represents the SEM across probes. **b.** Mean of normalized model-predicted responses (left) and actual neural responses (right) over time from stimulus onset, for top, middle, and bottom probes with respect to the ST location, during fixation (top; -500:-100 ms), and perisaccadic (bottom; -20:10 ms) windows. Values reported in Mean±SEM for models or neurons.



Supplementary Fig. 4. Decoding confusion in the direction same as and opposite to the saccade direction. Probe locations were more likely to be incorrectly decoded in the same direction as the saccade (blue), presaccadically (-100:0 ms) compared to perisaccadically (50:150 ms) (p = 1.21e-12). Probe locations were more likely to be incorrectly decoded in the opposite direction to the saccade (red), perisaccadically compared to presaccadically (p = 9.12e-06). Overlapping datapoints appear gray. Histogram in upper right show the distribution of differences.



Supplementary Fig. 5. Identifying bias-relevant STUs. a. Pictured are example kernels for two example probe locations over 1:200 ms delay, and the difference between them, over time. The AUC (yellow shaded area) represents the dissimilarity between kernels at two neighboring probes over time and delay. The process is repeated for STUs at 6 probe locations around the ST for all time and delay. **b.** Shows the map of bias index over STUs measured using difference between the full model AUC and the AUC with that STU removed. **c.** The cumulative distribution function of all the non-zero bias indices. Using the 90th percentile as a threshold, the STUs with an absolute bias index difference above 2.56 are defined as bias-relevant.



Supplementary Fig. 6. Model-predicted responses and bias-relevant STUs for MT and V4 neurons with RFs near or far from the ST. Color indicates model-predicted response, and black contours outline bias-relevant STUs, over time between stimulus presentation and saccade onset (y-axis) and time of response from stimulus onset (x-axis), for models of neurons recorded from MT (left) and V4 (right), for neurons with RFs near the ST (top), or far from the ST (bottom); *d* indicates the distance between the neurons' RF center and ST in dva.

Supplementary Table 1. Average spatial bias for different groups of neurons and saccade directions. Spatial bias in the fixation (-500:-300 ms) and perisaccadic (50:150 ms) time windows, for all ensembles (top row), ensembles divided by distance (*d*) between the RF center and the ST (middle row), and subdivided by saccade direction (bottom row). Negative spatial bias values indicate bias opposite the direction of the saccade (for both leftward and rightward saccades). Asterisks indicate the statistical significance of the difference in spatial bias between the fixation and perisaccadic windows (* $0.1 \leq p < 0.5$, ** $0.05 \leq p < 0.1$, *** $0.005 \leq p < 0.05$, **** $p < 0.005$). All mean±SEM. *p*-values were computed by one-sided Wilcoxon signed-rank test.

	spatial bias ** (15 ensembles, 447 neurons; $p = 0.083$)			
fixation	-0.024±0.060			
perisaccadic	-0.143±0.061			
distance between RF center and ST (dva)	***d* < 11** (4 ensembles, 94 neurons; $p = 0.625$)		***d* ≥ 11** **** (11 ensembles, 353 neurons; $p = 0.002$)	
fixation	-0.046±0.151		-0.015±0.063	
perisaccadic	0.055±0.130		-0.215±0.052	
saccade direction	**left** (2 ensembles, 21 neurons ; $p = 0.500$)	**right** (2 ensembles, 73 neurons; $p = 1.000$)	**left** *** (7 ensembles, 303 neurons; $p = 0.016$)	**right** * (4 ensembles, 50 neurons; $p = 0.250$)
fixation	-0.176±0.150	0.083±0.236	0.003±0.075	-0.047±0.106
perisaccadic	0.088±0.134	0.022±0.275	-0.260±0.061	-0.136±0.070

Supplementary References

1. Akbarian, A., Clark, K., Noudoost, B. & Nategh, N. A sensory memory to preserve visual representations across eye movements. *Nat Commun* **12**, (2021).

2. Brillinger, D. R. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol Cybern* **59**, 189–200 (1988).

3. Daley, D. J. & Vere-Jones, D. *An Introduction to the Theory of Point Processes*. (Springer New York, New York, NY, 2008). doi:10.1007/978-0-387-49835-5.